

Ontología para el control y recuperación de información onomástica en televisión

Por Jorge Caldera-Serrano y Rodrigo Sánchez-Jiménez

Resumen: Se muestra una propuesta de ontología creada para el control documental de información de personajes visualizados y referenciados en los servicios de documentación de las cadenas de televisión. Para ello se plantea el estado actual del uso de ontologías en el medio audiovisual, para posteriormente ir mostrando y explicando el desarrollo creado. Trabajo práctico donde se ofrece una solución rápida y sencilla para el control documental orientado específicamente a la recuperación de elementos onomásticos, por medio del cual se controlan y desarrollan las relaciones entre personajes. De esta forma se crea una auténtica red de personajes televisivos que puede ser explotada para mejorar los resultados durante el proceso de recuperación.

Palabras clave: Ontologías, Información audiovisual, Control onomástico, Recuperación de información, Televisión.

Title: Development and application of an ontology for onomastic information control and retrieval in television

Abstract: The development of an ontology for information control is proposed to manage visual images of and references to individuals in information management systems of television channels. Research has established precedents for the use of ontologies in the audiovisual field, providing a baseline for measuring and understanding ongoing developments. This practical approach offers a quick, straightforward solution to onomastic control in information retrieval, permitting control and development of relationships between individuals covered in the media. This creates a network of television personalities we can draw upon to improve the outcomes of retrieval processes.

Keywords: Ontologies, Audiovisual information, Named entities, Information retrieval, Televisión.

Caldera-Serrano, Jorge; Sánchez-Jiménez, Rodrigo. “Desarrollo y aplicación de una ontología para el control y recuperación de información onomástica en televisión”. En: *El profesional de la información*, 2007, enero-febrero, v. 17, n. 1, pp. 86-91.

DOI: 10.3145/epi.2008.ene.10



Jorge Caldera Serrano es profesor en la Facultad de Biblioteconomía y Documentación de la Universidad de Extremadura, donde imparte docencia sobre archivos de televisión, filmotecas y fuentes de información audiovisual. Es licenciado y doctor en Documentación por la Universidad de Salamanca y cuenta con la experiencia profesional de dos años en los Servicios Informativos de Televisión Española (Madrid) en el archivo audiovisual.



Rodrigo Sánchez Jiménez es profesor del Departamento de Biblioteconomía y Documentación de la Universidad Complutense de Madrid. Imparte docencia en materias de documentación informativa y técnicas de recuperación de información. Ha participado en varios proyectos de investigación relacionados con lenguajes documentales y documentación audiovisual y multimedia. Es autor de artículos sobre metadatos, documentación multimedia y recuperación de información.

Introducción

LAS CIENCIAS DE LA DOCUMENTACIÓN SIEMPRE SE HAN BASADO en el desarrollo de otros campos transversales para abrirse paso y especializarse. El hecho de ser una disciplina de apoyo y auxilio ha obligado a buscar mejores técnicas para la realización de nuestra labor, especialmente en lo relativo a la recuperación de información. La gestión de información en los medios

de comunicación es un fiel reflejo de la adaptación de herramientas documentales e informáticas al tratamiento documental.

Son muchos los intentos que se han llevado a cabo en esta línea, como la implementación de los tesauros desarrollados por *Antena 3TV* y *Televisión Española*, o tesauros facetados como el confeccionado por el grupo liderado por **Félix del Valle** y realizado para *Telecinco*, herramientas de clasificación, recu-

peración en texto libre, etc. Pero el gran crecimiento en cantidad y en exigencia de esta información hace que se sigan buscando nuevas vías para su control. Aunque los esfuerzos han sido amplios y ambiciosos, la realidad es que muchas de estas herramientas y técnicas aún no se adecuan a las necesidades de los periodistas ni a los requerimientos de los documentalistas.

El problema fundamental del análisis documental de imágenes en

movimiento es la descripción de acciones, ya que no son temas ni conceptos abstractos sino la representación de iconos que se modifican a lo largo del tiempo. Para resolver este aspecto se ha utilizado en ocasiones una combinación de tesauro y clasificación facetada, única forma viable de afrontar la descripción e identificación de acciones.

Las ontologías pueden constituir una buena solución a este problema dado su potencial y flexibilidad a la hora de interrelacionar conceptos, ideas y acciones, optimizando así la recuperación. La creación de herramientas ontológicas facilita la recuperación siempre y cuando la ontología creada haya sido planificada y organizada con criterios claros y pragmáticos. No obstante, una ventaja capital es su utilización para el intercambio de información entre empresas audiovisuales que utilicen una misma plataforma.

Las coordenadas de nuestro estudio se derivan de uno de los casos de aplicación más necesarios y evidentes: el control de personas físicas y jurídicas en el ámbito televisivo. No tratamos el ámbito temático por tratarse de un proceso diferente, con estructuras y relaciones ontológicas mucho más confusas y de compleja adaptación al modelo audiovisual.

Documentación en televisión

López de Quintana (2000) indica los principios fundamentales que debe regir la documentación en un medio televisivo; por un lado, la preservación del patrimonio documental generado por la cadena y,

por otro, integrar la documentación en sus actividades de producción y gestión para la creación de nuevos productos. La documentación en las empresas televisivas parte para **Hidalgo-Goyanes** (1999) de tres premisas fundamentales para el cumplimiento total de la misión y objetivos de los servicios: rapidez, eficacia y credibilidad.

López de Quintana (2000) realiza la división del organigrama atendiendo a diferentes parámetros tales como:

- Separación por tipo de actividad. Consiste en la división orgánica en la que se distingue la videoteca del centro de documentación, ya que la primera se responsabiliza del control físico del material audiovisual mientras que la segunda se encargaría de todo el proceso técnico.

- Atendiendo a la tipología documental. Suele ser habitual en los departamentos de documentación de las televisiones crear áreas de trabajo según la naturaleza del tipo documental, pudiendo encontrar secciones dedicadas a la documentación audiovisual, escrita, fotográfica y sonora.

- Por tipo de programas. Otra división que suele llevarse a cabo en las grandes cadenas televisivas con años de tradición es la separación en el organigrama documental dependiendo del tipo de programa sobre el cual se trabaja. La división se realiza entre programas de entretenimiento e informativos.

- Especialización temática y de técnicas documentales. Aunque sea cada vez menos habitual, en algunos centros el documentalista

está especializado en unas técnicas concretas, como el análisis o la difusión.

Queda claro que la gestión de la información en televisión va mucho más allá de la propia documentación audiovisual como se atestigua en **Caldera y Zapico** (2006), por lo que nos encontraremos en estos servicios con información fotográfica, sonora y textual.

No expondremos aquí de forma más pormenorizada los procesos de descripción en televisión ya que existen documentos más amplios que lo analizan (**Caldera; Nuño**, 2004), (**Fournial**, 1986). No obstante, sí es necesario subrayar la existencia de diferencias en el tratamiento y conservación de registros referenciales para programas informativos y de entretenimiento, consecuencia principalmente de los requerimientos de los usuarios y de los elementos formales intrínsecos de las piezas audiovisuales televisivas de ambas categorías.

El control actual de personas físicas y jurídicas en televisión

En el trabajo diario de las televisiones, la inspección de esta información básicamente se centra en un control de autoridad de los personajes aparecidos y referenciados en la información. Esto puede ser suficiente para muchas unidades informativas, pero no lo es siempre en los sistemas de información de las empresas televisivas. Pongamos un ejemplo representativo de este problema: ante la consulta “recursos de presidentes autonómicos de la década de los 80 en España”, su procesamiento no podría basarse exclusivamente en criterios onomásticos, sino que el documentalista tendría que acotar cronológicamente, consultar textos para identificar parcialmente cada uno de ellos y, por supuesto, guiarse por sus referencias culturales, informativas y formativas.

“Las ontologías cuentan con la ventaja de intercambiar la información entre empresas audiovisuales, además de ser flexibles interrelacionando conceptos, ideas y acciones”

No podemos dejar al criterio personal de los documentalistas este tipo de interrogación, sino que deben confeccionarse y prepararse herramientas de control que nos guíen en el momento de la consulta y recuperación. Aunque existen métodos de investigación documental para poder llegar a concluir con éxito la interrogación anterior, la diferencia fundamental es el tiempo de reacción con el que se cuenta en las televisiones. La existencia como mínimo de tres informativos y de una parrilla estable de programas, mucho de ellos en directo, provoca la necesidad de contestar en minutos a consultas en muchos casos extremadamente complicadas.

“La premura de tiempo en la resolución de las interrogaciones de los usuarios marca el ritmo y los métodos de trabajo. El futuro es el siguiente programa que se emite”

Estos puntos de acceso deben ser inspeccionados por medio de un control de autoridad, pero además pueden estar integrados en una ontología que aporte una mayor información, potenciando su recuperación por los diferentes elementos asociados a sus personas.

El trabajo con registros de programas informativos y de entretenimiento es bastante diferente debido a la diversa tipología documental. Para el caso que nos ocupa estimamos más factible desarrollar esfuerzos para potenciar una mejor y más coherente recuperación de personajes en los programas de entretenimiento, aunque también es útil y puede implementarse con mayor número de gestores de información para programas informativos.

Las relaciones entre los personajes de los programas de ocio tienen mayor importancia y utilidad. Al mismo tiempo es más factible llevar a cabo un control de autoridad exhaustivo debido a que el número de elementos que se integran semanalmente es más reducido. Piensen en los programas de concursos donde su número es importante y el de personajes es bastante limitado. Igualmente válidos son para los “reality show”, donde interesan los personajes protagonistas y las diferentes relaciones dentro y fuera del concurso. Por el contrario, en programas informativos el número de nuevos personajes es a veces desproporcionado, por lo que crear una ontología para controlarlos es factible, pero se traduce en mucho tiempo de dedicación por parte del documentalista.

El concepto “ontología” y la información audiovisual

A continuación se pretende aportar una revisión bibliográfica sobre el concepto de ontología, identificando aquellas definiciones que creemos más acertadas y allegadas al ámbito de lo audiovisual desde el punto de vista de la realización de nuestro trabajo.

Como labor de recopilación sobre este concepto indicamos el trabajo de **Corcho, Fernández y Gómez Pérez** (2003), que analizan las ontologías desde el concepto, los métodos y las herramientas de construcción. **García Jiménez** (2004) señala el objetivo último de las ontologías en cuanto a su desarrollo, definido como la mejora de la representación y recuperación de la información. **Noy y McGuinness** (2000) en su guía para la creación de ontologías, identifican sus objetivos principales: compartir la comprensión de estructuras entre personas o agentes software, permitir la reutilización del conocimiento perteneciente a un dominio, hacer explícitos los supuestos de un dominio, etc.

Sobre las definiciones aceptadas por la comunidad científica, la de **Gruber** (1993) es la más estandarizada: especificación explícita y formal de una conceptualización compartida. Este mismo autor (1993b) identifica los siguientes componentes:

- Conceptos: ideas que se intentan formalizar.
- Relaciones: representando la interacción entre los conceptos.
- Funciones: ofrecen un tipo concreto de relación, utilizados para la representación de objetivos determinados de un concepto.
- Axiomas: teoremas que identifican las relaciones existentes entre elementos.

Interesante es también la aportación de **Guerrero-Bote y Loza-**

Register for free at <https://www.scipedia.com> to download the version without the watermark

La información onomástica que podría definirse en la televisión es variada (**Caldera; Nuño, 2004**), no obstante indicamos las principales: personas físicas/referencias, entidades visionadas/referenciadas. La información que se extrae tanto de la banda de imagen como de la de sonido va en estancos separados, ya que a la hora de la recuperación no es idéntico el requerir una persona que se vea, una persona que hable o de la que trate la información (figura 1).

Puntos de acceso onomásticos	
Visionado	Referenciado
Personas físicas	Personas físicas
Personas jurídicas (entidades)	Personas jurídicas (entidades)

Figura 1. Esquema de puntos de acceso onomásticos

no (1999), quienes señalan que el dominio más cercano a la documentación son las ontologías confeccionadas desde el punto de vista terminológico (representación del conocimiento) e informativo (estructuración del almacenamiento de la información en las bases de datos). Será desde la perspectiva informativa desde la que abordemos y elaboremos nuestra propuesta de ontología.

Desconocemos la existencia de experiencias de diseño y creación de ontologías en las televisiones españolas o internacionales. La revisión de literatura científica no aporta resultados, al igual que los contactos directos con departamentos de documentación de diferentes cadenas estatales y autonómicas. Las empresas audiovisuales son compañías en las que el intercambio y difusión de información suele ser escaso, lo que no nos debe extrañar al ser corporaciones con idénticos fines, en lucha por la audiencia y los ingresos publicitarios.

Creación y desarrollo de la ontología

Se han de tener presentes una serie de elementos a la hora del diseño. Las ontologías deben ser claras, coherentes, adaptables, específicas y precisas. Claras para así comunicar de manera unívoca el significado. Coherentes para inferir de manera consistente entre definiciones. Adaptables para poder ofrecer nuevos usos. Específicas al detallar a nivel de conocimiento y precisas en su concepción y entendimiento.

“Las ontologías han de ser claras, coherentes, adaptables, específicas y precisas”

Nuestra labor se va a centrar en implementar una ontología en la que se especifica la estructura del almacenamiento de la información en una base de datos, mejorando así la identificación de las relaciones y optimizando el producto y resultado final ofrecido al usuario (Barchini, 2006)

Por todo ello, y con la necesidad de controlar los diferentes parámetros de personajes aparecidos en pantalla, se desarrolla en una primera fase un esquema que se expone en la figura 2.

Sobre cada uno de los personajes que aparecen en la parrilla televisiva se indicará al sistema tanto fecha y lugar de nacimiento como defunción en su caso, así como otras informaciones de carácter personal que deberán ser perfiladas en cada departamento de documentación atendiendo a las características de las solicitudes de los usuarios. Se definirán las posibles relaciones con entidades determinando las fechas, así como la responsabilidad del personaje en las mismas. Se intentará completar el mayor número de atributos en cada registro, aunque admitiendo que ocasionalmente algunos no se utilicen o puedan ser modificados.

No obstante, para una correcta gestión de esta información es necesaria la puesta en marcha y el se-

guimiento de un estricto control de autoridades, especificando así las diversas formas de denominaciones como variantes ortográficas, de tal manera que las interrogaciones por cada una de ellas siempre ofrezcan idénticos resultados. No queremos entrar en polémicas sobre esta cuestión, pero las normativas actuales de control de autoridades son excesivamente estrictas para la televisión. Además, el control en muchas cadenas es deficitario por lo que se debe comenzar realizando este esfuerzo normativo.

Se presenta en la figura 3 el modelo ontológico para gestionar la información onomástica en la base de datos de información audiovisual en una televisión.

Establecemos 5 clases primarias que representan los conceptos fundamentales para la descripción y recuperación de personajes. Los personajes constituyen la clase fundamental y son descritos en torno a características (*slots*) representados como rectángulos de borde azul para los atributos cuyo valor es tan sólo texto, como el nombre, los datos de control de autoridad (que incluirán variaciones ortográficas y diferentes versiones), y otros datos como las aficiones, su estado civil, etc. Respecto al control de autoridad, hay que decir que cada instancia que representa un personaje

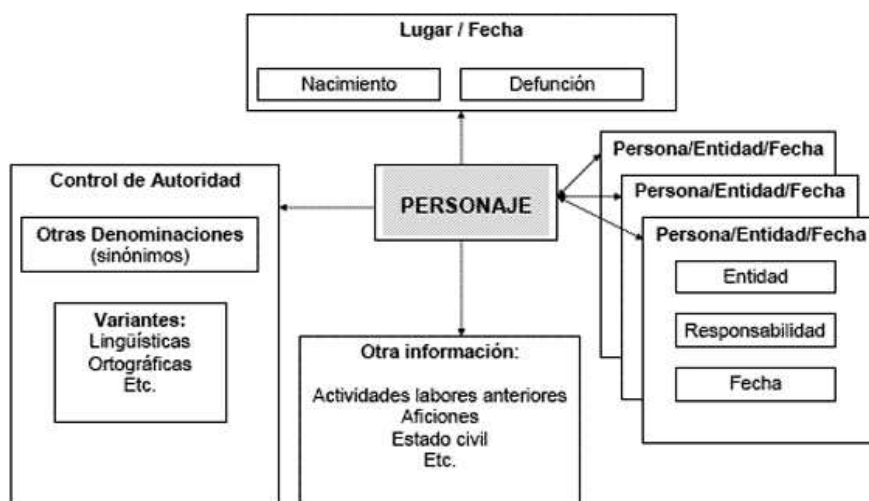


Figura 2. Control onomástico de personaje en TV

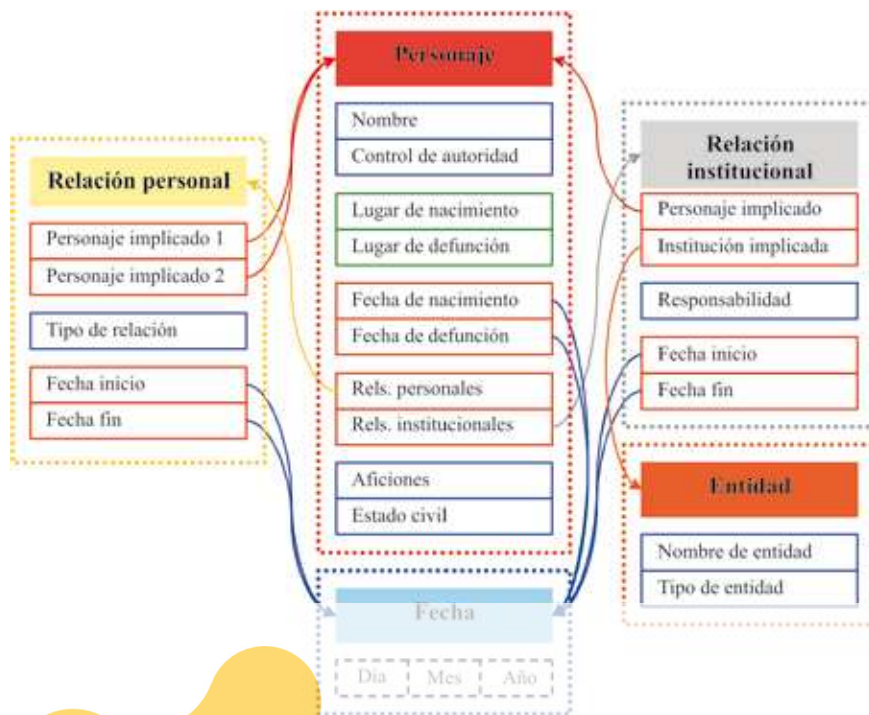


Figura 3. Estructura de ontología de personajes para TV

incluye todas las variantes de su nombre aparecidas hasta la fecha, y que una consulta cualquiera remitirá al mismo independientemente de la opción utilizada.

Las características marcadas como rectángulos de borde rojo definen vínculos con otras clases, otros conceptos de importancia para la recuperación, como relaciones personales e institucionales y las fechas. Por último, los lugares de nacimiento y defunción (rectángulos de borde verde) deberían indicarse utilizando otro recurso que permita el control del vocabulario, para lo que frecuentemente se cuenta con un tesoro geográfico. Esto permitiría mantener la consistencia en la descripción y facilitar la recuperación.

Las relaciones constituyen un concepto importante en sí mismas, por lo que se han representado como clases con sus propias características. Las fechas durante las cuáles tiene lugar la relación, así como el tipo específico de la misma, nos deberían servir para restringir los resultados de las búsquedas. Probablemente sería una buena idea

crear diferentes tipos de relaciones personales e institucionales como subclases de la clase principal, aunque este aspecto variará fuertemente en cada ocasión, por lo que nos limitamos a proponer una solución genérica. En el caso de las relaciones institucionales se vincula a una persona con una entidad, recordando las fechas y el tipo de responsabilidad. En cuanto a las relaciones personales se relaciona a dos personas diferentes indicando el tipo de relación, que podría ser laboral, amorosa, de familia, etc.

Las entidades pueden en realidad ser descritas mediante muchos más atributos que contribuirán a mejorar los resultados de las consultas, pero de nuevo nos limitamos a ofrecer unos mínimos que nos permiten hacernos una idea de los resultados en un entorno real. El tipo de entidad es de nuevo una característica que podría constituirse en un conjunto de subclases, aunque la profundidad y detalle con que se haga dependerá específicamente de la aplicación que se le quiera dar.

Por último, las fechas se han introducido como clases con un atri-

buto para los días, otro para los meses y otro para los años, cuyos valores deberían ser números enteros. El motivo fundamental es la necesidad de hacer cálculos sobre los valores indicados de manera que podamos resolver en una consulta criterios temporales con exactitud.

La aplicación de una ontología como la propuesta en la práctica pasa fundamentalmente por tres requisitos. En primer lugar es necesario valorar la necesidad de refinar el esquema de clases propuesto para adaptarlo a las necesidades reales de la descripción en un entorno concreto. Como ya anticipábamos podría ser necesario introducir una formalización más enriquecedora de los tipos de relaciones, así como de los tipos de entidades. En segundo lugar será preciso crear las instancias pertinentes para cada una de las clases (es decir, llevar a cabo la descripción de un número más o menos elevado de personajes para poder evaluar el sistema). Finalmente es necesario disponer de una forma de almacenar e interrogar la ontología.

Una vez se han establecido estas diferentes posibilidades a tener en cuenta. La primera de ellas es establecer una base de datos relacional sobre el esquema propuesto. Es una buena opción si no se dispone de herramientas específicas. Otra posibilidad, bastante más adecuada, es establecer un sistema de interrogación basado en *Sparql*, el lenguaje de consultas desarrollado por el W3C para *rdf* y *OWL*, los lenguajes que constituyen la espina dorsal de la web semántica. Esta solución es bastante más compleja, aunque permitirá explotar mejor la ontología. Por último se puede utilizar un razonador, un programa capaz de llevar a cabo inferencias que se derivan de la estructura de la ontología y de un conjunto de reglas establecidas mediante lógica descriptiva. Es la alternativa más compleja, aunque también la más potente.

“Las ontologías pueden constituir una opción excelente para la comprensión y utilización de las relaciones y procedencias de los personajes en televisión”

En nuestro caso, para probar la ontología propuesta utilizamos *Protégé*, un software de la *Universidad de Stanford* que permite desarrollarla y llevar a cabo consultas bastante complejas para explotarla. Durante el período de pruebas de la ontología, que se completó con la creación de varias decenas de personajes y relaciones, se observó cómo se podían resolver con precisión consultas como “recursos de presidentes autonómicos de la década de los 80 en España” o “personajes que mantuvieron una relación amorosa con empresarios de la construcción”. Sin embargo es necesario realizar un proceso de experimentación más profundo para llegar a resultados concluyentes.

Evidentemente el grado de éxito de la recuperación se deriva del tiempo y recursos empleados en el refinamiento de la ontología y su adaptación a las necesidades concretas del medio. Sin embargo, en conjunto, las ontologías proporcionan una forma de representar las complejas relaciones sociales y humanas de los personajes televisivos con un gran potencial de explotación.

Conclusiones

A modo de colofón se puede señalar que las televisiones actuales necesitan un mejor control terminológico de sus herramientas documentales, e igualmente una mejor adaptación a las cambiantes necesidades de los usuarios de los medios audiovisuales. Las herramientas cerradas y poco flexibles hacen que su viabilidad y potencialidad sean escasas. Y es en este escenario don-

de las ontologías muestran un desarrollo muy interesante.

En este trabajo se ha pretendido dar una utilidad parcial a las ontologías, solucionando un problema actual existente en el control y recuperación de los personajes visionados y referenciados en los programas de entretenimiento. Y decimos que parcial ya que la utilización de ontologías debe seguir desarrollándose en otras fases, incorporando elementos temáticos de material visualizado y referenciado para que realmente podamos encontrar por fin una herramienta ágil, ergonómica y adaptable a unas necesidades cambiantes.

Las ontologías pueden constituir una opción excelente para la comprensión y utilización de las relaciones y procedencias de cada uno de los personajes, lo que a veces no es nada sencillo en el contexto televisivo. Sin embargo, la viabilidad final del enfoque sólo puede ser probado y desarrollado sobre escenarios reales, con infraestructuras y datos de los que tan sólo las grandes corporaciones disponen. Por este motivo, nuestra propuesta pretende llamar la atención sobre las excelentes posibilidades que ofrecen las ontologías desde la perspectiva de la representación del dominio de los personajes televisivos, e incitar en lo posible a que se investigue en profundidad en el futuro.

Referencias

- Barchini, Graciela E.; Álvarez, Margarita M.; Herrera, Susana. “Sistemas de información: nuevos escenarios basados en ontologías”. En: *Revista de gestão da tecnologia e sistemas de informação*, 2006, v. 3, n. 1, pp. 3-18.
- Caldera-Serrano, Jorge; Nuño-Moral, M. V. *Diseño de una base de datos de imágenes para*

televisión. Gijón: Trea, 2004, Isbn 84-9704-100-3.

Caldera-Serrano, Jorge; Zapico-Alonso, Felipe. “Seen and heard: duality at the access point to television database”. En: *Aslib proceedings: news information perspective*, 2006, v. 58, n. 4, pp. 30-43.

Corcho, Óscar; Fernández-López, Mariano; Gómez-Pérez, Asunción. “Methodologies, tools, and languages for building ontologies. Where is their meeting point?”. En: *Data & knowledge engineering*, 2003, v. 46, n. 1, pp. 41-65.

Fournial, Catherine. “Análisis documental de imágenes en movimiento”. En: *Panorama de los archivos audiovisuales*. Madrid: Servicios de Publicaciones de RTVE, 1986, pp. 249-258, Isbn 84-85259-19-X.

García-Jiménez, Antonio. “Instrumentos de representación del conocimiento: tesauros versus ontologías”. En: *Anales de la documentación*, 2004, n. 7, pp. 79-95.

Gruber, Thomas R. “Toward principles for the design of ontologies used for knowledge sharing”. En: *Formal ontology in conceptual analysis and knowledge representation*. The Netherlands: Kluwer Academic Publisher, 1993.

Gruber, Thomas R. “A translation approach to portable ontology specifications”. En: *Knowledge acquisition*, 1993, v. 5, n. 2, pp. 199-220.

Guerrero-Bote, Vicente; Lozano-Tello, Adolfo. “Vínculos entre las ontologías y la biblioteconomía y documentación”. En: *La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de la información (Actas del IV Congreso ISKO-España Eoconsid '99)*. Granada: ISKO-Facultad de Biblioteconomía y Documentación, 1999, pp. 25-31.

Hidalgo-González, Paloma. “Documentación audiovisual”. En: García-Gutiérrez, Antonio (ed.). *Introducción a la documentación informativa y periodística*. Sevilla: MAD, 1999, pp. 473-485, Isbn 84-8311-460-7.

López-de-Quintana, Eugenio. “Documentación en televisión”. En: *Manual de documentación informativa*. Madrid: Cátedra, 2000, pp. 83-181, Isbn 84-376-1798-7.

Noy, Natalya F.; McGuinness, Deborah L. *Ontology development 101: a guide to creating your first ontology*, 2000.
<http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinnessabstract.html>

Jorge Caldera-Serrano, Departamento de Información y Comunicación, Universidad de Extremadura, Badajoz.
jcalser@alcazaba.unex.es

Rodrigo Sánchez-Jiménez, Departamento de Biblioteconomía y Documentación, Universidad Complutense de Madrid.
rsanchezj@ccinf.ucm.es